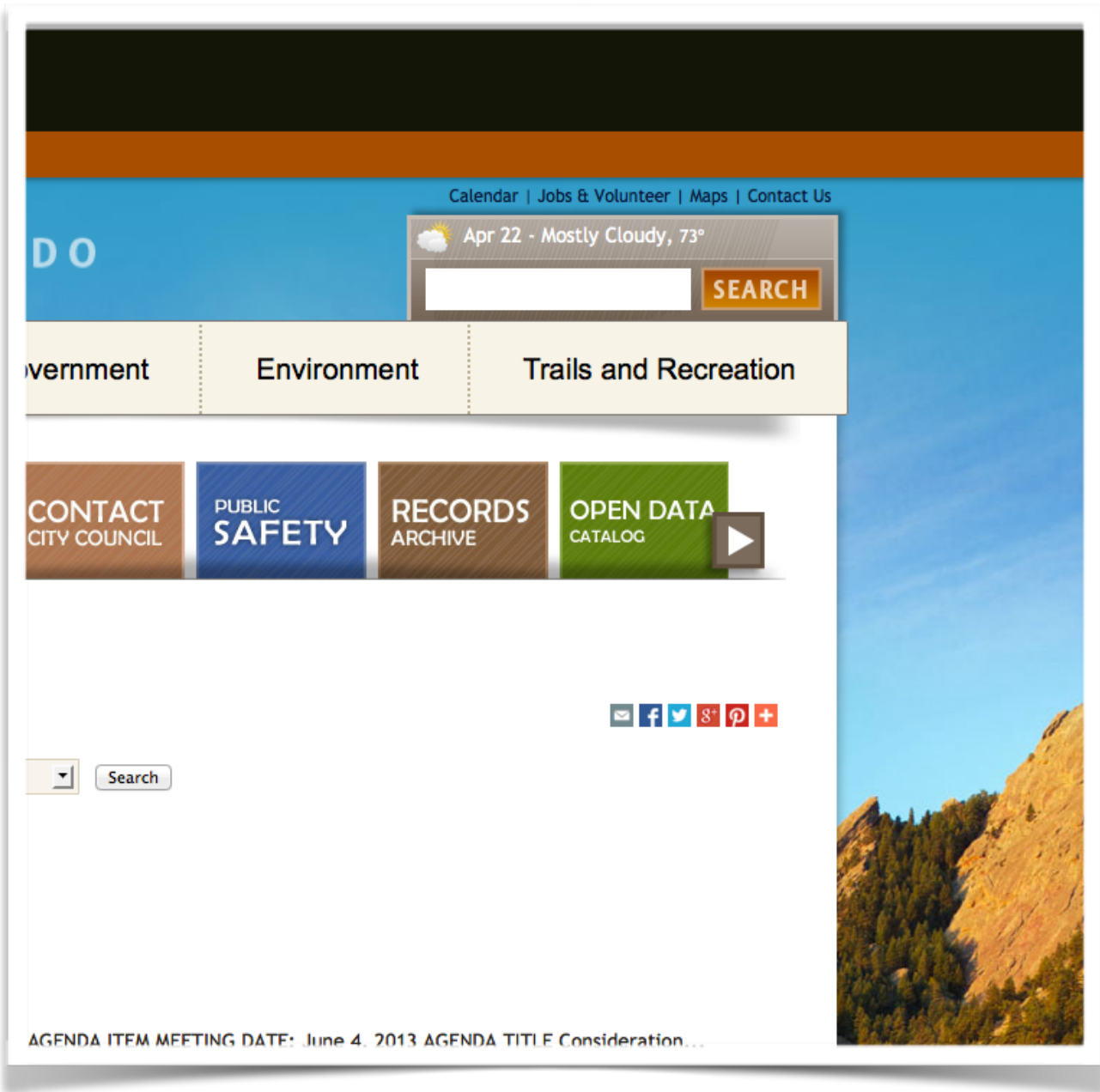# Integrated Search

What it is and how to use it effectively

This document serves as an overview of Integrated Search and outlines some tips to help you optimize your content for better findability through search.

TABLE OF CONTENTS

What Integrated Search Does and Does Not Do

Three Goals of Integrated Search

1. Improve speed of returning results.

   - The old Xpress search returned a page of results in, on average, six seconds.

   - Integrated search returns results in, on average, one second or less.

2. Broaden search to include new content in results.

   - We now index an multiple types of content including documents from Laserfiche (ECM), FAQs from InquireBoulder (CRM), along with pages, news, links and files from Xpress. We also index open data and job listings.

3. Improve search accuracy.

   1. Integrated Search greatly improves accuracy as compared to the previous built-in Xpress search. We continue to take steps as a group to maintain and improve that accuracy.

What is Elasticsearch?

At the core of Integrated Search is an open source indexer called Elasticsearch. Elasticsearch maintains the indexes in its own highly optimized database and will be indexing content on a regular basis.

Xpress content is indexed every five minutes while Laserfiche documents are indexed every 24 hours because they change less often. FAQs, open data and job listings are indexed every 12 hours.

What Does the Search Index

It's important to know what content is indexed because that's the content that gets searched when someone enters a word or phrase in the **search** field. What gets indexed varies by content type.

- **InquireBoulder FAQs** - Title, keywords and body
- **Xpress Web Pages** - URL, category, title, summary, body, tags, related links
- **Xpress Links** - Title, summary, tags
- **Xpress Files** - Title, summary, tags, document content
- **Laserfiche Files** - Title, summary, category, department, document type, document content

What Does the Search Not Index

Integrated Search currently does not index content included via short codes on pages in Xpress, including content in includes files or iFrames. While it does index metadata related to MP3 files, it does not index the actual audio content.

It does not index the content of external files (for instance, a PDF we might link to on the State of Colorado's website), but it does index the metadata provided for that link.

Integrated Search also does not index content that is not considered publicly available. Only Laserfiche documents that are shared through the public portal are indexed. Only Xpress pages, files and links with a status of "On" that are published and not expired are indexed.

If you delete a file or a page, it may take up to five minutes before it is removed from the search indexer.

Accuracy

What are some accuracy considerations when using Integrated Search?

Accuracy is a continuous process, not a static goal. The accuracy of Integrated Search very much depends on the content being indexed. Consider the following examples:

- Integrated Search will not find a page on "employment" when the user searches for the term "jobs" unless the employment page includes the word "jobs" somewhere in the content or metadata.
- Integrated Search will not find a page that has all of its content included via an iFrame unless the relevant keywords and phrases are included in the title, summary or tags for that page.

Part of the job of a content manager is to not only create the content but to ensure that the content includes the appropriate metadata to make it findable via search. This applies just as much to your content's findability via external search engines like Google and Bing.

10 Tips for improving Integrated Search accuracy

1. Make sure your content is in text format.
   - If you're embedding a large JPG info graphic on your page in lieu of actual text, your content will not be findable.
2. Pay attention to Readability Statistics which are displayed below the body content section of every page when you're in edit mode.
   - In general, content that reads at an 8th grade level is the goal. This means you'll be using words and phrases in your content that the average user is also likely to use in their searches.
3. Provide summaries for your content. Include key phrases and relevant words.

4. Tag your content!

    • Use tags to reinforce findability

5. Use meaningful tags.

    • Tagging something with the phrase "parksandrecreationannualduckrace" isn't going to be very helpful for search. Instead, tag the same content with key words and phrases like "parks and recreation", "rec", "duck race".

    • There is no need to include words that are already in a phrase tag. Using the above example, if you tag content "parks and recreation", then you don't need to use the tags "park" or "recreation", but you will want to use "rec".

    • Don't use tags that have no relevance to your content. This dilutes search results with irrelevant content and causes users to perceive the search as inaccurate.

6. With the exception of Xpress, tags/keywords are simply a way of relating words and phrases to content to help with findability. Keep in mind that Xpress tags are also used as a way of grouping content together (for use in page short codes, for instance).

7. Create relevant titles for your content that use key words or phrases.

8. Monitor the Top 10 Searches - you can view the last 7 days and the last 30 days on the admin page when you first log in to Xpress.

    • Each search term is linked directly to the search result page for that term.

    • If you see a term that applies to your content, be sure to check the search result page to make sure the right content is appearing.

    • If a page or link or file isn't appearing, and you feel it should, review your content to be sure you've applied these tips.

    • If a page or link or file *is* appearing and you feel it shouldn't, confer with the manager of that content (if it isn't you) and make changes as appropriate.

9. Unpublish, delete and remove old and irrelevant content! Keep in mind that all "deleted" content is still retained in the Xpress database so it will still meet document retention policies.

10. If you are still having problems with inaccurate results for a particular search term, or can't seem to optimize a piece of content for search, contact Leslie or Ron for help.

Remember that curating your content is an ongoing, continuous process. We're counting on you to work with us to make your content easily findable and create a good experience for the public and staff.